

EXTENDED REPORT

Microsurgical lab testing is a reliable method for assessing ophthalmology residents' surgical skills

J B Taylor, G Binenbaum, P Tapino, N J Volpe

Br J Ophthalmol 2007;**91**:1691–1694. doi: 10.1136/bjo.2007.123083

See end of article for authors' affiliations

Correspondence to: N J Volpe, Scheie Eye Institute, 51 North 39th Street, Philadelphia, PA 19104, USA; nickvolp@mail.med.upenn.edu

Accepted 2 June 2007
Published Online First
25 June 2007

Background: Formal assessment of clinical competencies is necessary to ensure that all residents are acquiring important skills and, in the United States, will soon become a requirement for residency programme accreditation by the Accreditation Council for Graduate Medical Education (ACGME). The Eye Surgical Skills Assessment Test (ESSAT), a laboratory-based surgical skills obstacle course, was developed in response to the need for improved tools for the assessment of surgical skills during residency. The ESSAT has previously been shown to have face and content validity, and in this study we sought to determine its inter-rater reliability and, to some extent, its construct validity.

Methods: Twenty-seven content experts (residency programme directors and faculty members involved with resident surgical training) watched videos of a junior resident and senior resident completing the three ESSAT stations (skin suturing, muscle recession, and phacoemulsification: wound construction & suturing technique) and completed assessment forms, both task-specific checklists and a global rating scale of performance.

Results: The ESSAT showed strong inter-rater reliability for determining whether a resident "passed" a threshold of competency at each station for both the checklists and global rating scale. In addition, for each station, the senior resident was consistently rated above a "passing" threshold using either assessment form, whereas the junior resident was more often rated below (94% vs 30% passing on completed forms).

Conclusion: These results, along with the findings of our face and content validity analysis, support the reliability and validity of the ESSAT, and indicate that it could be a useful tool for improving the assessment of surgical skill during residency. The ESSAT is a tool that all residency programmes could implement as a part of their ophthalmic surgical curriculum and competency assessment, and may be useful to set a threshold of competence that all residents would need to achieve prior to entering the operating room.

New assessment tools are needed to improve the process of teaching and evaluating residents in core competencies. In the United States, this need has become a mandate, as the Accreditation Council for Graduate Medical Education (ACGME) has set forth a timeline by which all residency programmes, in order to maintain their accreditation, must develop and integrate new tools for teaching and evaluating residents in six core competencies.¹ Surgical skills was added as a seventh competency by the American Board of Ophthalmology.² The Eye Surgical Skills Assessment Test (ESSAT) was developed both as a response to these mandates and to the need for more objective and structured methods of assessing residents' surgical skills. In ophthalmology, several new surgical skills assessment tools have recently been developed: in addition to the wet lab-based ESSAT, programmes will have in their armamentarium procedure specific evaluation forms (which many programmes have already been using), assessment of OR performance with videos as well as specific forms^{3–4} and simulation technology.^{5–6}

Modelled after the Objective Structured Assessment of Technical Skills (OSATS), a laboratory-based surgical skills-assessment test developed⁷ and validated⁸ by researchers in the field of general surgery, the ESSAT is made up of three simulated surgical tasks that the resident is required to complete in the microsurgical laboratory. These tasks are (1) skin suturing, (2) muscle recession and (3) phacoemulsification: a wound construction & suturing technique. The resident's performance may be observed live or on videotape by a surgical educator who completes a task-specific checklist as well as a global rating scale of performance for each task.

The ESSAT offers the controlled setting of the microsurgical laboratory for residents to learn and be assessed in a

standardised fashion. In addition, the ESSAT takes skills assessment and basic competency determination out of the operating room, where patient risks become involved. We previously established that the ESSAT has face and content validity by surveying experts in the field and incorporating their suggestions for improving the ESSAT.⁹ To ensure that the ESSAT has the test characteristics needed of a good assessment tool (ie, validity, reliability), we set forth in this study to also establish the inter-rater reliability and, to a limited degree, the construct validity of the ESSAT, particularly for the purpose of establishing a threshold of basic skills competency that all residents must achieve in order to enter the operating room.

MATERIALS AND METHODS

Two US ophthalmology residents, a junior resident (in the first year of ophthalmology training) and a senior resident (in the third and final year of ophthalmology residency training), agreed to participate and to be videotaped completing each of the three ESSAT tasks. Residents were given 15 min to complete each station. Ophthalmology residency programme directors were invited to participate via the American Academy of Ophthalmology's email distribution list for programme directors. Some programme directors forwarded the email to other surgical educators. Those who volunteered to be raters were mailed an explanation of their task along with a CD, which contained three of the six recorded video segments. The six possible video segments were (1) junior resident completing skin suturing station; (2) senior resident completing skin suturing station; (3) junior resident completing muscle recession station.

Abbreviations: ACGME, Accreditation Council for Graduate Medical Education; ESSAT, Eye Surgical Skills Assessment Test

sion; (4) senior resident completing muscle recession; (5) junior resident completing phacoemulsification: wound construction & suturing technique; (6) senior resident completing phacoemulsification: wound construction & suturing technique. Each volunteer received two videos of one station (completed by both the junior and the senior resident) and one additional video segment of another station. Volunteers were asked to complete assessment forms, both task-specific checklists (a step by step list of the procedure broken down with a forced choice of yes or no for adequate performance) and a global rating scale (typical 5-point Likert scale rating the resident on various aspects of a well-done procedure) of performance, for each video segment that they watched. As an attempt to create a "blinded" observer, the participants were not informed of the level of training of the residents they were watching in each video.

The data were analysed for inter-observer reliability and for consistency in determining whether a resident has reached a certain threshold of competency, with the belief that such a threshold could be used to determine when a resident is adequately prepared, in terms of their basic surgical skills, to enter the operating room. The "passing" threshold was set at $\geq 70\%$ of items correct on the checklist for each station and a score of ≥ 3 for each item on the 5-point global rating scale.

RESULTS

Fifty-three residency programme directors and surgical educators originally volunteered to participate. Of these, 27 (51%) returned completed assessment forms. For Station 1: Skin Suturing, 19 experts watched the video of the junior resident. Ninety-four per cent of raters (16/17, two raters were not included due to incomplete data on their assessment forms) gave the resident a passing score on the checklist, and 42% (8/19) gave the resident a passing threshold score on the global rating scale. Six expert surgeons watched the senior resident complete the skin suturing station and 100% (6/6) rated this resident above the threshold for both the checklist and global rating scale. For Station 2: Muscle Recession, six experts watched the junior resident. None of these experts (0/6) gave the junior resident a passing score on the checklist, and 17% (1/6) gave a passing score on the global rating scale. At Station 2, the senior resident was given a passing score for the checklist by 100% (6/6) and for the global rating scale by 83% (5/6) of the raters who watched this video. Eighteen raters watched the junior resident video for Station 3: Phacoemulsification—Wound Construction & Suturing Technique. Of note, this was the one station performance where, as a result of the 15-min time limit, the junior resident did not complete the last two items on the checklist. For the checklist, one rater was eliminated because he did not fully complete the checklist form. This left 17 raters, none of whom (0/17) gave the resident a passing score for checklist. For the global rating scale, again none (0/18) of the raters gave the resident a passing score. Finally, 20 raters watched the video of the senior resident completing Station 3. Ninety per cent (18/20) gave a passing score for the checklist, and 95% (19/20) gave a passing score for the global rating scale.

Figure 1 summarises these data in terms of the consistency (inter-rater reliability) with which raters scored the resident in each video clip above or below the threshold level of competency. For each of the checklists, at least 90% of the raters were consistent with one another regarding whether the resident passed the threshold of competence. For the global rating scale, at least 80% of the raters had consistent ratings for each resident at each task, except for the scores for the junior resident at the skin suturing task, which showed only 58% consistency among raters.

Figures 2–4 compare the ratings of the junior resident to the senior resident. The ratings show that two of the ESSAT tasks (muscle recession, phacoemulsification: wound construction & suturing technique) consistently discriminate between the two residents at different levels of training (construct validity). Combining all of the assessment forms, the senior resident was rated above the "passing" threshold 94% (60/64 forms) of the time, whereas the junior resident was only rated above the threshold 30% (25/83 forms) of the time.

DISCUSSION

All new assessment tools should be shown to have certain test characteristics, namely reliability and validity, before being put into general use. For the assessment of resident surgical skills, the ESSAT thus far seems to satisfy these criteria. In a previous study, we established the face and content validity of the ESSAT and now have data suggesting that this surgical skills assessment test has good inter-rater reliability. This study also shows, to a limited degree, evidence for construct validity. In this context, construct validity is used to mean the ability to discriminate between residents at different levels of training.

Having now demonstrated these test characteristics, the ESSAT may be adopted by residency programmes as a new, structured and more objective method for teaching and assessing residents' surgical skills. This tool may be useful to ensure that residents reach a basic level of competency prior to entering the operating room where teaching is less controlled, and patient risks cannot be eliminated. In addition, by facilitating the use of the microsurgical laboratory and emphasising basic surgical skills, the ESSAT may be able to improve the overall process of early surgical education in ophthalmology residency and become an important part of the resident's surgical competency portfolio. For residency programmes in the United States, the ESSAT may be implemented as one step towards satisfying the demands of the ACGME and maintaining accreditation.

The primary aim of this study was to assess the inter-rater reliability of the ESSAT. While the responses of experts were not identical, they were consistent in their overall determination of competency, which was set at a threshold of $\geq 70\%$ of checklist items marked as correct and $\geq 3/5$ on each of the global rating scale items. For each of the checklists, at least 90% of the raters were consistent with one another regarding whether the resident passed the threshold of competence. For the global rating scale, at least 80% of raters gave consistent scores with the exception of when they were rating the skin suturing station. At this station, 40% (8/19) of raters scored the junior resident above the passing threshold, and 60% (11/19) scored this resident below. One factor that may have contributed to this lack of inter-rater reliability is that more than half of the raters who "failed" this resident also watched the video of the senior resident at this station. In fact, this video clip would have been just prior in sequence to the clip of the junior resident on the CDs that were mailed. None of those raters who "passed" this resident had also received the video clip of the senior resident. Most likely, those raters who received both clips were more critical of the junior resident because they had just finished watching the clip of the more skilled senior resident. The inconsistency of the ratings exposes the fact that the global rating scale is not wholly objective. These inconsistencies may be diminished once a rater has seen several residents perform the ESSAT and has developed a good sense of where different residents should fall on the range of the scale. Interestingly, there was no inconsistency of rating skin suturing using the checklist. This may partially be due to the fact that the forced choice, binary (correct or incorrect/not done) checklists are less qualitative than the global rating scale.

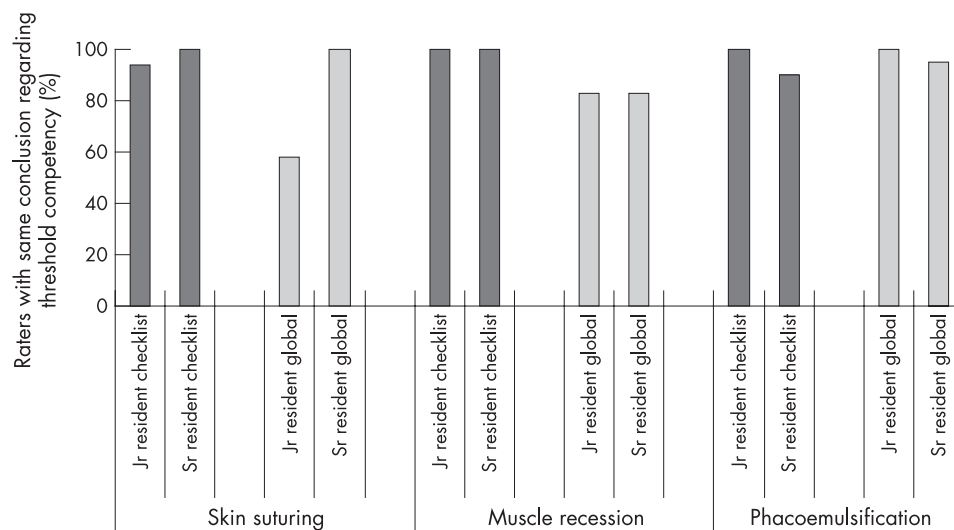


Figure 1 Percentage of raters who came to the same conclusion as one another (inter-rater reliability) regarding the resident's performance of each task as either passing or not passing the threshold competency (defined as $\geq 70\%$ of items correct on the checklist for each station and $\geq 3/5$ on each item on the global rating scale).

As this was the first trial run of the ESSAT, the video quality was less than perfect and there were several segments that had problems with image stability and focus. Several raters commented on this issue, which may have contributed to some of the variability in scores on any particular video segment. A few raters who submitted incomplete assessment forms noted that they did so because they could not clearly see some portion of the task. With time and experience, the ESSAT video techniques will be refined, perhaps resulting in a further increase in the reliability of ratings.

The ESSAT appears to have the ability to distinguish between a junior and a senior resident (construct validity). Overall, the senior resident was rated above the "passing" threshold 94% of the time, whereas the junior resident was only rated above the threshold 30% of the time. This finding offers some evidence of construct validity. The individual who was likely to be a more competent surgeon, the senior resident, did better on the ESSAT than the individual who was less likely to have well-developed surgical skills. The ESSAT was not as good at discriminating the residents at the skin suturing station. This may be because of a ceiling effect for this relatively straightforward task. Even the novice junior resident was able to "pass" the threshold of competency for this task. As only two residents were being compared, further testing is clearly needed to assess the ESSAT's ability to discriminate between residents. Further

testing on more residents will also help determine whether the ESSAT can pick up even more subtle differences, such as that between a resident in the first year of training and one in the second year of training or between an average and a particularly talented senior resident.

The number of raters who were mailed each different video segment was not equivalent. As a result, we have much more data regarding station three (phacoemulsification: wound construction & suturing technique) for both residents and station one (skin suturing) for the junior resident. More data need to be collected to make sure that the inter-rater reliability and construct validity of the other three segments remain strong with larger numbers of raters. In addition, as we collect more data, we will be able to refine the "passing" thresholds to ensure that they are set at a level that is appropriate to require residents to reach before beginning training in the operating room. It should be emphasised that the ESSAT is not intended to be a stressful test that will prevent struggling residents from beginning their training in the operating room. Rather, the ESSAT is a form of quality assurance for residents to guarantee that they have adequate exposure to surgical techniques prior to beginning real-life surgical training. Emphasis will be placed on providing constructive feedback through the specific items marked incorrect on the checklists and on facilitated and guided practice in the microsurgical laboratory.

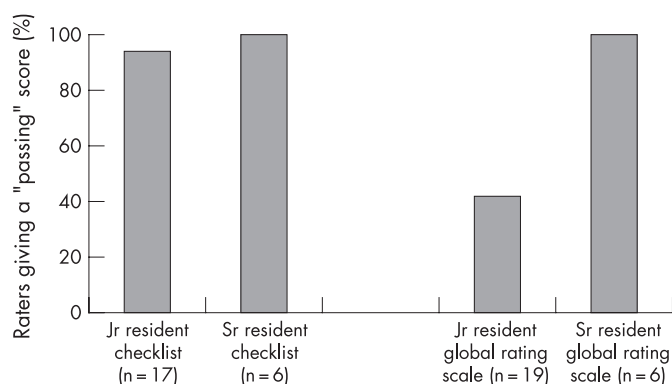


Figure 2 Percentage of raters who determined the resident's performance to be above the set passing threshold of competence ($\geq 70\%$ of items correct on the checklist and $\geq 3/5$ on each item on the global rating scale) for Station 1: Skin Suturing.

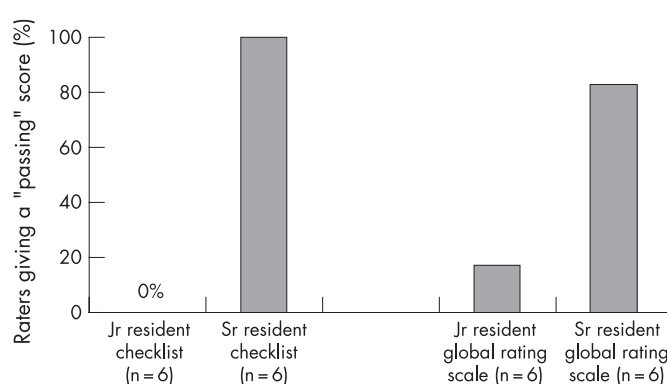


Figure 3 Percentage of raters who determined the resident's performance to be above the set passing threshold of competence ($\geq 70\%$ of items correct on the checklist and $\geq 3/5$ on each item on the global rating scale) for Station 2: Muscle Recession.

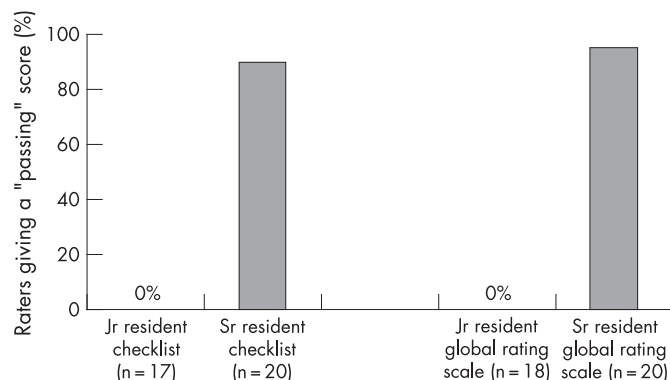


Figure 4 Percentage of raters who determined the resident's performance to be above the set passing threshold of competence ($\geq 70\%$ of items correct on the checklist and $\geq 3/5$ on each item on the global rating scale) for Station 3: Phacoemulsification—Wound Construction & Suturing Technique.

Ensuring the surgical competency of residents is a critical component of every ophthalmology residency programme. The traditional forms of surgical skills assessment, unstructured summative faculty evaluations written at the end of a rotation and faculty meetings with discussion of residents' abilities, are inadequate. We have now collected evidence that the ESSAT is a reliable and valid new assessment tool. Complemented by other new assessment methods that will evaluate residents' skills in the operating room^{3,4} and with simulation technologies,^{5,6} the ESSAT could help guarantee that all residents achieve surgical competence during their residency training.

Authors' affiliations

J B Taylor, Department of Ophthalmology, University of California—San Francisco, CA, USA

P Tapino, N J Volpe, Department of Ophthalmology, University of Pennsylvania School of Medicine, Scheie Eye Institute, Philadelphia, PA, USA

G Binenbaum, Department of Ophthalmology, Children's Hospital of Philadelphia, Philadelphia, PA, USA

JBT was supported by a Doris Duke Fellowship, and GB was supported by a Heed Fellowship and Society of Heed Fellows Fellowship.

Competing interests: None.

Material in this manuscript was presented at the Association for University Professors of Ophthalmology Annual Meeting, February 2007, the Scheie Eye Institute 132nd Anniversary Meeting, May 2006 and the American Academy of Ophthalmology Annual Meeting, November 2006.

REFERENCES

- 1 Accreditation Council for Graduate Medical Education (ACGME) general competencies. <http://www.acgme.org> (accessed 2005).
- 2 Lee AG, Carter KD. Managing the new mandate in resident education: a blueprint for translating a national mandate into local compliance. *Ophthalmology* 2004;**111**:1807–12.
- 3 Cremers SL, Ciolino JB, Ferrufino-Ponce ZK, et al. Objective Assessment of Skills in Intraocular Surgery (OASIS). *Ophthalmology* 2005;**112**:1236–41.
- 4 Cremers SL, Lora AN, Ferrufino-Ponce ZK. Global Rating Assessment of Skills in Intraocular Surgery (GRASIS). *Ophthalmology* 2005;**112**:1655–60.
- 5 Henderson BA, Neaman A, Kim BH, et al. Virtual training tool. *Ophthalmology* 2006;**113**:1058–9.
- 6 Webster R, Sassani J, Shenk R, et al. Simulating the continuous curvilinear capsulorhexis procedure during cataract surgery on the EYESI system. *Stud Health Technol Inform* 2005;**111**:592–5.
- 7 Martin JA, Regehr G, Reznick R, et al. Objective Structured Assessment of Technical Skill (OSATS) for surgical residents. *Br J Surg* 1997;**84**:273–8.
- 8 Faulkner H, Regehr G, Martin J, et al. Validation of an objective structured assessment of technical skill for surgical residents. *Acad Med* 1996;**71**:1363–5.
- 9 Fisher JB, Binenbaum G, Tapino P, et al. Development and face and content validity of an eye surgical skills assessment test (ESSAT) for ophthalmology residents. *Ophthalmology* 2006;**113**:2364–70.

Take advantage of BMJ Journals' remarkable catalogue of titles with Related Collections

No busy professional has time to browse through all pertinent journals to find relevant articles, but with Related Collections you no longer have to. Follow the "Related Collections" link from any article and use the "Show Collections from other Journals" to expand your search across all BMJ Journals. Or simply follow the "Browse by topic" link on the home page. By setting up your own collections and receiving email alerts every time an article is added to your chosen area, you can build up your own significant body of knowledge.